Synthesis and CD Studies of an 88-residue Peptide Containing the Main Receptor Binding Site of HTLV-I SU-glycoprotein

MICHAEL GOETZ[1], SERGE GEOFFRE[1], BERNARD BUSETTA[1], CLAUDE MANIGAND[1], CLAUDE NESPOULOUS[2] DANIÈLE LONDOS-GAGLIARDI[3], BERNARD GUILLEMAIN[3], and MICHEL HOSPITAL[1]

[1] Unité de Biophysique Structurale, CNRS, Université de Bordeaux, France
[2] Laboratoire d'Études de Protéines, INRA, Jouy-en-Josas, France
[3] INSERM U328, Institut Bergonié, Bordeaux, France

Abstract: Essential HTLV-I biological functions, like host-cell receptor recognition, depend on the structural motives on the surface glycoprotein gp46. We defined a peptide of 88 amino acids [Arg^{147}-Leu^{234}] corresponding to the central part of the protein sequence, where major neutralizing epitopes are localized. After evaluating the feasibility of its chemical synthesis, the chosen sequence was realized using the stepwise solid-phase methodology. Multiple chromatographic purification steps were required to obtain a sample suitable for structural analysis. Correct folding was supported by strong binding of monooclonal antibodies, recognizing known exposed immunodominant regions. Circular dichroism studies confirmed a non-random conformation of at least 70–80% of the synthetic peptide. Investigation of the 3D-structure of the synthetic peptide will provide useful information for future vaccine and drug-design strategies. © 1997 European Peptide Society and John Wiley & Sons, Ltd.

The human retrovirus HTLV-1 (human T-cell leukemia virus type I) has been shown to be the etiological agent of adult T-cell leukemia (ATL) [1,2] and/or tropical spastic paraparesis, also termed HTLV-I associated myelopathy (TSP/HAM) [3,4]. The envelope of the virus, like those of other retroviruses, is composed of two glycoproteins (a surface unit gp46 non-covalently linked to the transmembrane unit gp21) originating from a glycosylated precursor encoded by the HTLV-I *env* gene. The gp46 plays an important role in viral infection (viral absorption and mediation of cell-penetration by recognition of specific receptors expressed on the surface of the target cells, cytopathogenic effects) and elicits specific antibodies, some among them being neutralizing [5,6]. The low genetic heterogeneity observed suggests genetic pressure to conserve the envelope protein structure and function. Indeed the HTLV-I envelope is much less variable in its amino-acid sequence than that of HIV-I; the extent of variability being 3.1% at most in the cosmopolitan subtype [7],

Address for correspondence: Dr Michael Goetz, Unité de Biophysique Structurale, CNRS–Université de Bordeaux, 351 Cours de la Libération, 33405 Talence cédex, France.

whereas the HIV-I envelope protein is under high functional constraints [9,10], so that elucidation of its three-dimensional structure should contribute to the understanding of the mechanisms involved in receptor recognition.

The 3D-structure of only one envelope glycoprotein complex, that of influenza A haemagglutinin, has been resolved on atomic scale by X-ray diffraction studies [11–13]. Different molecular modelling studies suggest that retroviral glycoproteins conserve the general organization of haemagglutinin [14], but may present structures that differ in the central portions of their external subunits [15]. This central part, a globular domain called the head, corresponds to a very immunogenic region and presents the major neutralizable epitopes [16]. Suggestion of a potential receptor binding site in this region is also supported by alignment of the head domains of several oncovirus glycoproteins [14–15]. Basing ourselves on these results, we defined different peptides of moderate length after molecular modelling studies, corresponding to structural domains on gp46 that we supposed being sufficiently stable 'on their own' to get realized by chemical synthesis. We retained two slightly different sequences of HTLV-I cosmopolitan strain MT-2, [Phe^{161}-Pro^{243}] and [Arg^{147}-Leu^{234}]. These model peptides should be representative for the binding site of gp46 to its principal host-cell receptor.

Peptides were synthesized using Merrifield's solid-phase peptide synthesis (SPPS) methodology

[17]. Our first approach was to check the feasibility of the defined sequences [Phe^{161}-Pro^{243}] and [Arg^{147}-Leu^{234}]. To minimize costs we started the syntheses on a small scale (1 mmol). Progress in synthesis was controlled at several stages of chain-assembly by amino acid analysis (AAA) of hydrolysed samples of peptide-resin. The distribution of residues made it possible to check the elongation of the sequence. The first resin sample was taken after incorporation of Arg^{231} at position 12 (the first arginine in the sequence [Phe^{161}-Pro^{243}]). AAA clearly showed its presence in the hydrolysate and thus indicated the absence of extensive chain termination at this point of the synthesis. A second sample, taken after incorporation of Glu^{219} (position 24), showed a decrease in the number of peptide chains still growing (ratio T/S less than the 3/3 expected). The third hydrolysate (after Glu^{201}) confirmed our concern: no lysine (Lys^{207}, Lys^{209}) could be detected. That means incoming amino acids were no longer coupled to the peptide chain.

In SPPS glutamine residues are known to favour interchain association by means of H-bonding involving the δ-amide functions of their side chains. Regarding the primary amino acid composition of the model peptide, we considered [Gln^{215}-Leu- Thr-Leu-Gln^{219}] as a 'difficult sequence', which might be responsible for the premature interruption of chain elongation. The homologous pentapeptide in the parent murine retrovirus (moloney murine leukemia virus) corresponds to [Asp-Gly-Ala-Tyr-His]. We
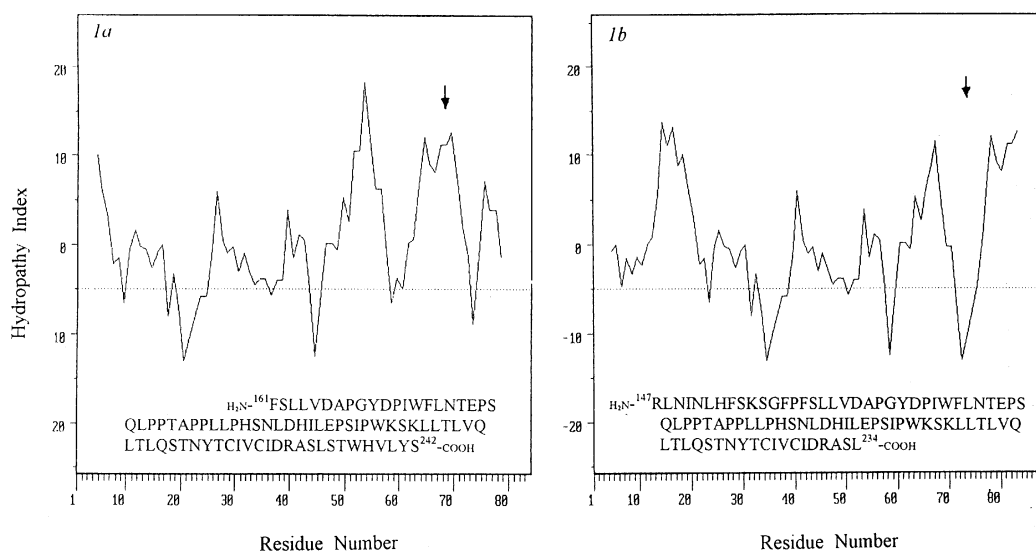

Figure 1 Hydropathy index as determined by SOAP/PC-GENE and primary sequences of the 82-residue peptide [Phe^{161}– Pro^{243}] (a) and the 88-residue peptide [Arg^{147}–Leu^{234}] (b). The arrows indicate the 15th residue, counting from the C-terminus.
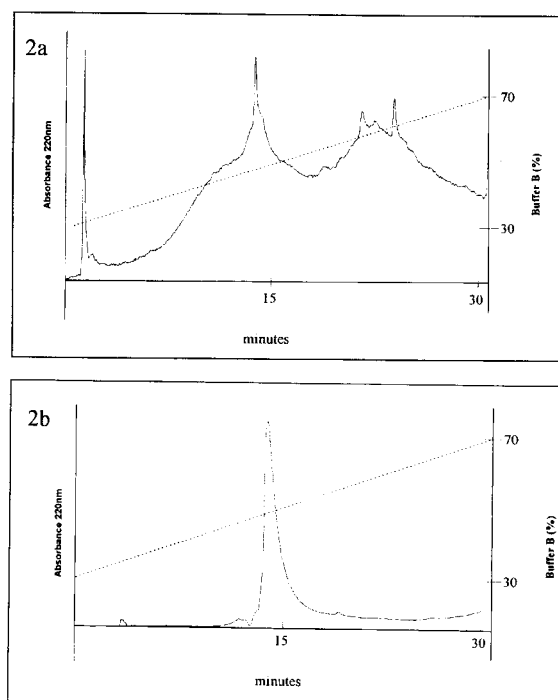
Figure 2 Analytical Reversed Phase HPLC of the crude synthetic peptide (a) and the pure product (b). The elution was made with 0.1% aqueous TFA (buffer A) and 0.08% TFA in acetronitirle (buffer B) at a flow rate of 1 ml/min, where buffer B varied in a 30 min linear gradient form 30 to 70%.

replaced the primary sequence respectively, to see whether the problems could be explained this way. As AAA showed again a premature termination in chain elongation, the problem had to be located elsewhere.

Another well-known problem in SPPS is a physical one. The solid support is composed of functionalized polystyrene resin beads, containing pores where the peptide chain gets assembled. Somewhere between the 10th and 15th residue the growing peptide 'leaves' the pores and gets in direct contact with the reaction solvents, a much less hydrophobic environment. The hydrophobic character of the ensuing amino acid sequence can be decisive for the success or failure of the synthesis. In the case of sequence $[Phe^{161}–Pro^{234}]$, the hydropathy profile (Figure 1(a)) shows the unfavourable conditions. Sequence $[Arg^{147}–Leu^{234}]$, which differs in only a nine amino acid shift towards the protein's N-terminus, favours a hydrophilic region after the tenth residue of this peptide (Figure 1(b)). During a small-scale synthesis of $[Arg^{147}–Leu^{234}]$, we could now follow the chain elongation up to the 88th residue, using again AAA, and confirm our hypothesis.

The finally retained model-sequence $[Arg^{147}–Leu^{234}]$ was synthesized on a standard scale (0.5 mmol/g). To maintain high coupling yields, double
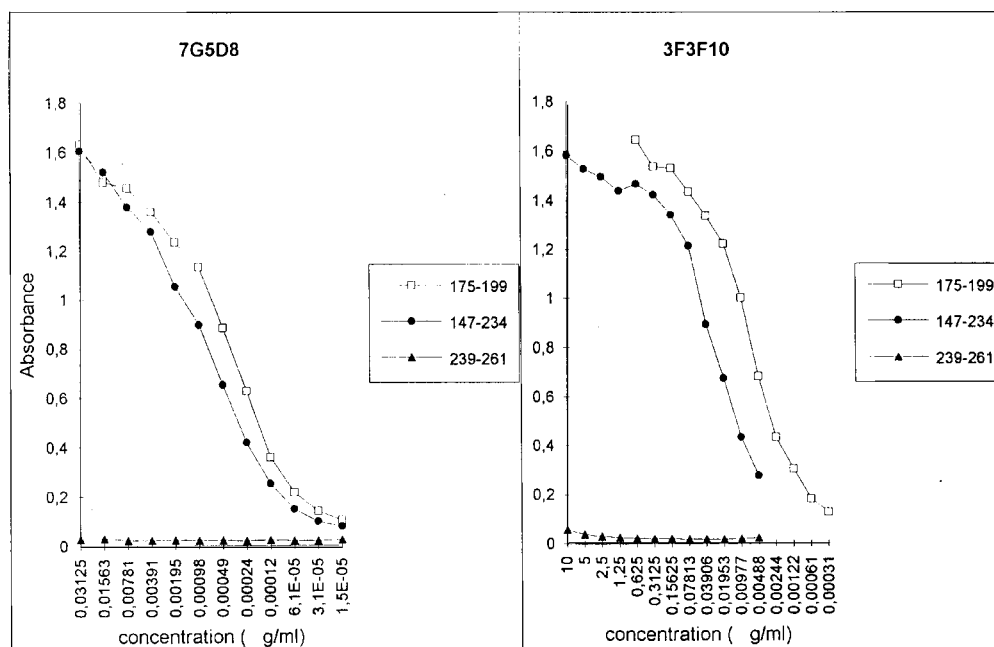

Figure 3 Binding of monoclonal antibodies *7G5D8* and *3F3F10* (serial log^2 dilutions) to synthetic peptides: 175–199 positive control (□), 147–234, 88-residues model peptide (●) and 239–261 negative control (▲). Both mAb recognized the model peptide $[Arg^{147}–Leu^{234}]$ with high titres: 7.5 ng/ml for *7G5D8* and 30 ng/ml for *3F3F10*.
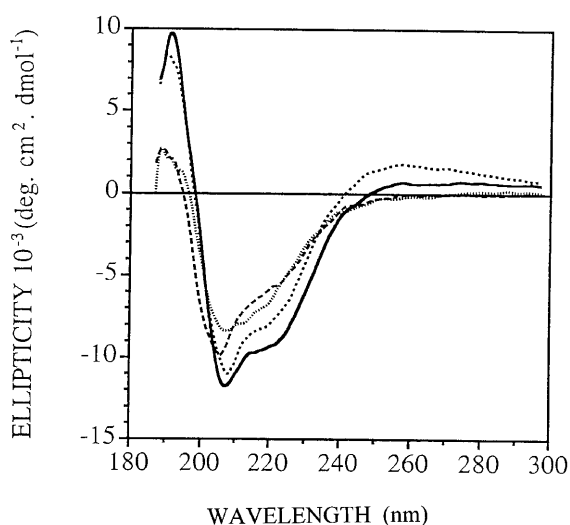
Figure 4 CD spectra of the synthetic peptide in various concentrations of acetonitrile in water: $\cdots\cdots$ 5%; $----$ 19%; - - - - - 40%; ———— 60%.

couplings were performed for all residues. After the final cleavage of the peptide from the solid support, the crude product was submitted to a gel filtration on a Sephadex G50 column in order to separate the target product from the truncated sequences, resulting of the blocking of the unreacted amino group by acetylation during the synthesis. The fractions were lyophilized, identified by AAA, SDS–PAGE electrophoresis and analysed by RP-HPLC on an analytical column. Automated Edman degradation
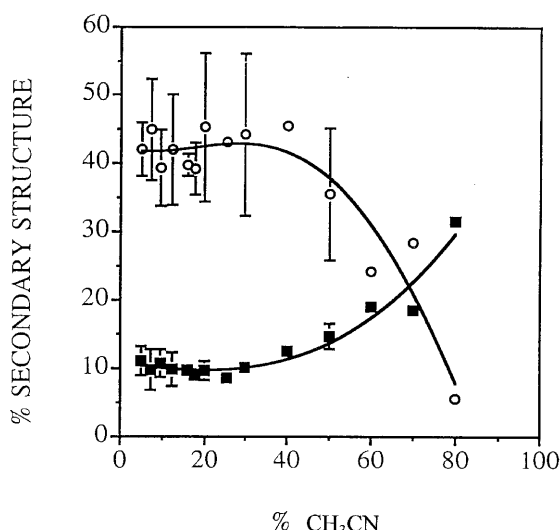
Figure 5 Effect of acetonitrile on the secondary structure of the synthetic peptide in aqueous mixtures: α-helix (■), β-sheet (GEEMPTYCIR;). Except for 40, 60, 70 and 80%, each point is the mean value $\pm$ SD of three measurements.

of the peptide's N-terminus was performed in order to verify the correct sequence. Unambiguous identification of the first six amino acids confirmed the success of peptide synthesis. The next steps of purification, semi-preparative RP-HPLC, led to a large increase in purity of the synthetic peptide (Figure 2(a) and 2(b)). The desired product has been judged sufficiently pure for structural analysis experiments.

To confirm monoclonal antibody (mAb) binding capacity of the model peptide [Arg[147]–Leu[234]], we tested its recognition by two of our mAb directed against the principal neutralizable region of gp46 [18]: *7G5D8* recognizes a linear epitope between amino acids Leu[183] and Leu[191], whereas *3F3F10* recognizes a conformational epitope inside the sequence Phe[175]–Leu[199] [19]. The strong bindings of both mAb suggest the ability of [Arg[147]–Leu[234]] to adopt the three-dimensional folding necessary to the correct exposure of the immunodominant region Phe[175]–Ile[199] of gp46 (see Figure 3 for binding profiles and mean titers).

The model peptide corresponds to an exposed domain of an extracellular protein, but nevertheless it proved to be insoluble in pure water and aqueous buffers. In order to identify secondary structure elements in solution and to optimize solubilization conditions, we used circular dichroism (CD) spectroscopy. Testing different organic solvents, including trifluoroethanol, we finally chose acetonitrile to dissolve the peptide, since this solvent was used in post-synthesis purification. Peptide [Arg[147]–Leu[234]] must first be dissolved in 50% (v/v) acetonitrile, but the organic solvent concentration can subsequently be reduced to as little as 5% (v/v). Figure 4 shows the CD spectra of [Arg[147]–Leu[234]] in four different concentrations of acetonitrile. The spectra have similar general shapes, characterized by a positive band near 190 nm. Few variations in the relative percentages of the peptide's different secondary structure elements seem to occur in aqueous solutions with increasing proportions of acetonitrile (Figure 5). Increasing the concentration of organic solvent leads to an increase in the amplitude of the positive and negative bands (very strong in the case of the positive band). Analysis of the spectra by the least squares method shows that the structure of our model peptide is composed of 50% β-sheet and very little α-helix at low concentrations of acetronitrile; the rest corresponds chiefly to random coil structure. Above 50% acetonitrile, the peptide seems to adopt a more α-helical structure, the random coil contribution increasing slightly.

In conclusion, we have realized the total chemical synthesis and purification of an 88-residue peptide that we consider as a representative model of HTLV-I's main receptor binding site. The strong binding titers of two monoclonal antibodies directed against the immunodominant Phe[175]–Ile[199] region of gp46 suggested the correct presentation of the peptide's central sequence; CD-studies confirmed a non-random conformation of at least two-thirds of the peptide chain over a wide range of solvent composition. The determination of the model peptide's three-dimensional structure may help to elucidate the role of retroviral envelope glycoproteins in host-cell infection. This could provide useful information for future vaccine- and drug-design strategies.

EXPERIMENTAL PART

Solid-phase Peptide Synthesis

The chemical synthesis of the 88 amino acids sequence was performed automatically on an Applied Biosystems Peptide Synthesizer, Model 431A, using the t-Boc strategy as modified by Applied Biosystems. N-α-protected leucine linked to 4-(oxy-methyl)phenyl-acetaminomethyl-copoly(styrene -1% divinyl-benzene) resin and other N-α-Boc-protected amino acids were from Novabiochem (Switzerland). DCC, HOBt, 1,2-ethanedithiol, piperidine, mercaptoethanol, DMSO from Sigma (Sigma-Chime, France); DMF, dicholoromethane (DCM), N-methyl-pyrrolidone (NMP), TFA, diisopropylethylamine (DIEA), methanol, acetic acid and acetonitrile were purchased from SDS (SDS-PEYPIN, France); HF was obtained from UCAR (Belgium). The t-Boc-N-α-amino acids were protected as follows: cyclohexyl ester for aspartic and glutamic acids, 2-bromo-carbobenzoxyl for tyrosine, p-toluenesulphonyl for arginine, dinitrophenyl for histidine, formyl for tryptophan, benzyl for serine and threonine, acetamidomethyl for cysteine and trityl for glutamine [20].

Elongation was done on Boc-Leu-PAM-resin (0.5 mmol; loading of starting resin: 0.4 mmol/g) according to the following operational cycle protocol: deblocking with 50% TFA/DCM, washing with DCM, neutralization with 10% DIEA/DCM, washing with DCM, then coupling in NMP with 2 mmol preformed HOBt esters of side-chain protected Boc amino acids. DMSO and DIEA were added to increase the coupling efficiency by helping to break self-aggregation of peptides. To facilitate post-synthesis purification a capping-step was included after the second coupling using acetic anhydride. Coupling efficiencies were monitored using the quantitative ninhydrin test. After coupling of the 39th residue, half of the peptide-resin was removed form the reaction vessel before continuing the synthesis; the remaining half was stored in DMF at 4 °C for elongation to be continued.

Cleavage of the Peptide from the Resin

After coupling of the final amino acid, the dinitrophenyl groups of histidine residues were removed by thiolysis. The peptide–resin was stirred in a solution of DMF containing 20% mercaptoethanol and 10% DIEA for 1 h at room temperature. After draining and thoroughly washing the resin with DMF, ethanol and finally DCM, the N-terminal Boc group was removed in the above-mentioned fashion. To deformylate the protected Trp residues prior to cleavage with HF, the peptide–resin was treated with 10% piperidin in DMF for 2 h at 0 °C, washed with DMF, DCM and ethanol and dried overnight in a vacuum desiccator. The peptide was cleaved from the resin by classical high HF treatment [21], with anisol and ethanedithiol as scavengers. Good splitting of the tosyl protection-group of arginine was achieved by allowing a reaction time of 90 minutes at 0 °C. HF was then rapidly evaporated *in vacuo*, peptide and resin were triturated with 50% acetic acid. After dilution with water, the cleaved peptide was lyophilized.

Amino Acid Analysis (AAA) and Sequencing

The peptide was hydrolysed in 6 M HCl (150 °C, 1 h) using a Millipore-Waters Pico-Tag work station; amino acids were derivatized to their phenylthiohydantoin conjugates before RP-HPLC, using the Pico-Tag column. Amino acid sequences were confirmed by comparison with calibration analyses using 100 pmol amino acid standards supplied by the manufacturer. AAA of the purified 88-residue peptide gave: Ala 3.11 (3), Arg 1.89 (2), Asx 9.19 (9), Gly 2.09 (2), Glx 5.15 (5), His 2.91 (3), Ile 5.94 (6), Leu 16.36 (16), Lys 3.08 (3), Phe 3.82 (4), Pro 11.40 (11), Ser 8.77 (9), Thr 5.92 (6), Tyr 1.87 (2), Val 3.02 (3).

Automated Edman degradation of the peptide's N-terminus up to the sixth amino acid was performed on an Applied Biosystems 475A Protein Sequencer for further control.

Size-exclusion Chromatography

The crude peptide was loaded to a 2.5×120 cm^2 column of Sephadex G50 fine (Pharmacia Biotech SA, France) and eluted with 50% acetic acid. Elution of the product was monitored at 220 and 280 nm. The fractions were analysed by reverse-phase HPLC and those containing the desired product pooled and lyophilized.

Analytical HPLC and Semi-preparative HPLC Purification

The lyophilized product was dissolved in 0.1% aqueous TFA and centrifuged. The supernatant was analysed by HPLC on a Waters 600E System Controller and Waters 996 Photodiode Array Detector using a C$_4$-Delta Pak column (3.9×150 mm^2, 300 Å, 5 μm). The elution was made with 0.1% aqueous TFA ($=$ buffer A) and acetonitrile (containing 0.08% TFA $=$ buffer B) at a flow rate of 1ml/min, where acetonitrile varied in a 30 min linear gradient form 30 to 80%. Semi-preparative HPLC was performed using a C$_4$-Delta Pak column (7.8×300 mm^2, 300 Å, 15 μm) in two consecutive purification steps: elution with a linear gradient (the same as on analytical scale) was followed by elution under isocratic conditions (70/30 buffer A/B).

ELISA

Synthetic peptides were coated in carbonate-bicarbonate buffer at pH 9.6 overnight at 4 °C in 96 well plates (Maxisorb, Nunc). The wells were saturated with 2% bovine serum albumin for 1 h at 37 °C, then incubated with serial log^2 dilutions of mAb for 2 h at 37 °C. After washing with PBS containing 0.05% Tween 20, wells were incubated with peroxydase-labelled anti-mouse IgG (H $+$ L) (Sigma) for 45 min at 37 °C. Enzymatic activities were measured with ABTS as substrate by a plate reader at 405 nm; the titers given correspond to the mAb concentration (μg/ml) at which the absorbance is equal to half of the maximum value. Synthetic peptides were used as positive [Phe^{175}–Ile^{199}] and negative [Val^{239}–Ala^{261}] controls.

Circular Dichroism

CD spectra were recorded at 1 nm intervals over the wavelength range 180–300 nm, using a Mark V Jobin-Yvon dichrograph. The dichrograph was calibrated using iso-androsterone (Roussel-Uclaf) in dioxane and camphorsulfonic acid (Sigma) in water. The optical rotation was checked with cytochrome C (horse heart, Fluka), lysozyme (chicken egg white, Merck) and L-lactate dehydrogenase (rabbit muscle, Fluka). Peptide samples were first dissolved in 50% organic solvent and then diluted 10-fold with appropriate mixtures containing decreasing percentages of solvent. CD measurements were performed at room temperature in a 0.2 cm path-length cell with peptide concentrations about 50 μg/ml. Five reproducible spectra were collected for each sample. The net spectrum of each protein was obtained by subtracting the base line obtained with the corresponding solvent only. The CD data were expressed as the mean residue ellipticity and the spectra analysed following the method of Yang *et al.* [22] for the determination of the contribution of the different secondary structures. The method was controlled using the proteins that are mentioned above. Peptide concentrations were determined spectrophotometrically using an extinction coefficient of 14,000 M^{-1}cm^{-1} at 280 nm, calculated form Gill and von Hippel [23].

REFERENCES

1. B. J. Poiesz, F. W. Ruscetti, A. F. Gazdar, P. A. Bunn, J. D. Minna and R. C. Gallo (1980). Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc. Natl. Acad. Sci. USA 77*, 7415–7419.
2. I. Miyoshi, I. Kubonishi, S. Yoshimoto, T. Akagi, Y Ohtsuki, Y. Shiraishi, K. Nagata and Y. Hinuma (1981). Type C virus particles in a cord T-cell line derived by co-cultivating normal human cord leukocytes and human leukaemic T-cells *Nature 294*, 770–771.
3. A. Gessain, F. Barin, J. C. Vernant, O. Gout, L. Maurs, A. Calender and G. De Thé (1985). Antibodies to human T-lymphotropic virus type 1 in patients with tropical spastic paraparesis. *Lancet 2*, 407–410.
4. M. Osame, K. Usuku, S. Izumo, N. Ijichi, H. Amitani, A. Igata, M. Matsumoto and M. Tara (1986). HTLV-I associated myelopathy, a new clinical entity. *Lancet 1*, 1031-1032.

5. N. Yamamoto, J. Schneider, Y. Koyanagi, Y. Hinuma and G. Hunsmann (1983). Adult T-cell leukemia (ATL) virus-specific antibodies in ATL patients and healthy virus carriers. *Int. J. Cancer 32*, 281–287.

6. Y. Inoue, H. Kuroda, H. Shiraki, H. Sato and Y. Maeda (1992). Neutralizing activity of human antibodies against the structural protein of human T-cell lympho-tropic virus type 1. *Int. J. Cancer 52*, 877–880.

7. A. Gessain, R. C. Gallo and G. Franchini (1992). Low degree of HTLV-I genetic drift *in vivo* as a means of monitoring viral transmission and movement of an-cient human population. *J. Virol. 66/4*, 2288–2295.

8. B. R. Starcich, B. H. Hahn, G. M. Shaw, P. D. McNeely, S. Modrow, H. Wolf, E. S. Parks, W. P. Parks, S. F. Josephs, R. C. Gallo and F. Wong-Staal (1986). Identification and characterisation of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell 45*, 637–648.

9. C. Pique, T. Tursz and M. C. Dokhelar (1990). Muta-tions introduced along the HTLV-I envelope gene result in a non-functional protein: A basis for envelope conservation? *EMBO J. 7*, 4243–4248.

10. L. Delamarre, C. Pique, D. Pham, T. Tursz and M. C. Dokhelar (1994). Identification of functional regions in the human T-cell leukemia virus type 1 Su glycopro-tein. *J. Virol. 68/6*, 3544–3549.

11. I. A. Wilson, J. J. Skehel and D. C. Wiley (1981). Structure of the haemagglutinin membrane glycopro-tein of influenza virus at 3 Å resolution. *Nature 289*, 366–373.

12. P. A. Bullough, F. M. Hughson, J. J. Skehel and D. C. Wiley (1994). Structure of the influenza haemaggluti-nin at the pH of membrane fusion. *Nature 371*, 37–43.

13. T. Bizebard, B. Gigant, P. Rigolet, B. Rasmussen, O. Diat, P. Boesecke, S. A. Wharton, J. J. Skehel and M. Knossow (1995). Structure of influenza virus haemag-glutinin complexed with a neutralizing antibody. *Nature 376*, 92–94.

14. B. Busetta (1989). The use of folding patterns in a multisolution approach of the all-β protein three dimensional structure. A β-roll structure is predicted in the retroviral glycoprotein. *Biochem. Biophys. Acta 998*, 301–309.

15. I. Callebaut, D. Portetelle, A Burny and J. P. Mornon (1994). Identification of functional sites on bovine leukemia virus envelope glycoproteins using structural and immunological data. *Eur. J. Biochem. 222*, 405–414.

16. E. Baba, M. Nakamura, Y. Tanaka, M. Kuroki, Y. Itoyama, S. Nakano and Y. Niho (1993). Multiple neutralizing B-cell epitopes of human T-cell leukemia virus type 1 identified by human monoclonal anti-bodies. *J. Immunol. 151*, 1013–1024.

17. G. Barany and R. B. Merrifield in: The Peptides, E. Gross and J. Meienhofer, Eds., pp. 1–284, Academic Press, New York 1980.

18. P. Horal, W. W. Hall, B. Svennerholm, J. Lycke, S. Jeansson, L. Rymo, M. H. Kaplan and A. Vahlne (1991). Identification of type specific linear epitopes in the glycoproteins gp46 and gp21 of human T-cell leukemia virus type I and type II using synthetic peptides. *Proc. Natl Acad. Sci. USA 88*, 5754–5758.

19. E. Edouard, E. Legrand, T. Astier-Gin, R. Dalibart, S. Geoffre, P. Dalbon, B. Guillemain and D. Londos-Gagliardi (1994). Characterization of monoclonal anti-bodies directed against the gp46 of human T-cell leukemia virus type I. *Leukemia 8*, S60–S64.

20. P. Sieber and B. Riniker (1991). Protection of carbox-amide functions by the trityl residue. Application to peptide synthesis. *Tetrahedron Lett. 32/6*, 739–742.

21. M. Mutter, G. C. Tuchscherer, C. Millert, K. H. Altmann, R. I. Carey, D. F. Wyss, A. M. Labhart and J. E. Rivier (1992). Template-assembled synthetic proteins with 4-helix bundle topology. Total chemical synthesis and conformational studies. *J. Am. Chem. Soc. 114*, 1463–1470.

22. J. T. Yang, C. S. C. Wu and H. M. Martinez (1986). Calculation of protein conformation from circular dichroism. *Methods Enzymol. 130*, 208–269.

23. S. C. Gill and P. H. von Hippel (1989). Calculation of protein extinction coefficients from amino acid se-quence data. *Anal. Biochem. 182*, 319–326